

AD\_\_\_\_\_

Award Number: W81XWH-~~€~~ ~~€~~ ~~€~~ ~~€~~ H

TITLE: 0Z0^){ { ^Ë ã^Á^æ oOæ &^!À&æ / Á Ææ Æ ^!ææ •

PRINCIPAL INVESTIGATOR: Óscar J. Alpa

CONTRACTING ORGANIZATION: University of U[ ~ @!} Âœä!} æ  
Š • Ä \* ^ ^ • ÊÖÖÄ €€ JÁ

REPORT DATE: June 20FF

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
1. REPORT DATE (DD-MM-YYYY) 01-06-2011		2. REPORT TYPE Annual		3. DATES COVERED (From - To) 1 JUN 2010 - 31 MAY 2011	
4. TITLE AND SUBTITLE  A Genome-Wide Breast Cancer Scan in African Americans				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-08-1-0383	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)  Christopher Haiman   E-Mail: haiman@usc.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  University of Southern California Los Angeles, CA 90089				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Genome-wide association studies (GWAS) of breast cancer have been completed among populations of European ancestry, and several regions have been identified that appear to contribute susceptibility to this cancer. Recent data suggests that not all risk alleles for common cancers will be revealed however by studies limited to Whites of European ancestry, and that similar efforts in other racial and ethnic populations will be needed to identify the full spectrum of common risk alleles that contribute to disease risk in the population. To identify genetic risk alleles for breast cancer risk among African American women we have performing a well-powered whole-genome association scan. For this project we have established a collaborative network of investigators whose careers have been dedicated to studying breast cancer in minority populations who have contributed samples and covariates from each of their respective studies to identify genetic variants that contribute to risk of breast cancer in this minority population. We have completed a GWAS of >1.1 SNPs in >3000 African American breast cancer cases and >2,700 controls. With these data we have validated and improved upon markers of risk at the known breast cancer risk regions that better characterize their contribution to breast cancer risk in women of African ancestry. In collaboration with GWAS in populations of European ancestry we have also revealed a novel risk locus for estrogen receptor negative breast cancer.					
15. SUBJECT TERMS  Breast Cancer, Genome-Wide Association Study, African Americans					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT  UU	18. NUMBER OF PAGES  22	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

## Table of Contents

	<u>Page</u>
Introduction.....	1
Body.....	2-9
Key Research Accomplishments.....	10
Reportable Outcomes.....	11
Conclusion.....	12
References.....	13-18
Appendices.....	19

## Introduction

Genome-wide association studies (GWAS) of breast cancer have been completed among populations of European ancestry, and several regions have been identified that appear to contribute susceptibility to this cancer. Recent data suggests that not all risk alleles for common cancers will be revealed however by studies limited to Whites of European ancestry, and that similar efforts in other racial and ethnic populations will be needed to identify the full spectrum of common risk alleles that contribute to disease risk in the population. To identify genetic risk alleles for breast cancer risk among African American women we have performing a well-powered whole-genome association scan. For this project we have established a collaborative network of investigators whose careers have been dedicated to studying breast cancer in minority populations who have contributed samples and covariates from each of their respective studies to identify genetic variants that contribute to risk of breast cancer in this minority population. We have completed a GWAS of >1.1 SNPs in >3000 African American breast cancer cases and >2,700 controls. With these data we have validated and improved upon markers of risk at the known breast cancer risk regions that better characterize their contribution to breast cancer risk in women of African ancestry. In collaboration with GWAS in populations of European ancestry we have also revealed a novel risk locus for estrogen receptor negative breast cancer.

## BODY

*The Specific Aim of this application is to identify genetic risk alleles for breast cancer among African American women by performing a well-powered genome-wide association study (GWAS).* For this project, I have established a network of leaders in the breast cancer research community with long-standing interests in breast cancer research in African Americans, all of whom have existing case-control studies of breast cancer in the U.S. Funding for the genotyping of samples from the MEC, CARE, WCHS, SFBC and BCFR studies is covered by this DOD-BCRP grant. Genotyping was conducted using the Illumina Infinium 1M. The genotyping of the other studies has been provided by a number of other sources. Stage 1 of the GWAS included 9 epidemiological studies of invasive breast cancer among African American women, which comprise a total of 3,153 cases and 2,831 controls. Details of the participating studies, genotyping and statistical analysis of the GWAS data have been provided in previous progress reports. Here we present results for two specific analyses: 1) fine-mapping of the known breast cancer risk loci, and 2) a meta-analysis of GWAS for estrogen receptor negative breast cancer.

### **Fine-Mapping of Breast Cancer Susceptibility Loci Characterizes Genetic Risk in African Americans**

We tested common genetic variation at the breast cancer risk loci identified in women of European and Asian descent in the stage 1 African American breast cancer sample to identify markers of risk that are relevant to this population. More specifically, we examined the index variants and conducted fine-mapping of the locus to both improve the current set of risk markers in African Americans as well as to identify new risk variants for breast cancer. We then applied this information to model breast cancer risk in African American women in attempt to characterize the spectrum of genetic risk in this population defined by common variants at the known risk loci.

We tested the 19 validated breast cancer risk variants (referred as “index variants”) at 1p11, 2q35, 3p24, 5p12, 5q11, 6q25, 8q24, 9p21, 9q31, 10p15, 10q21, 10q22, 10q26, 11p15, 11q13, 14q24, 16q12, 17q23 and 19p13 in models adjusted for age, study, global ancestry (the first 10 eigenvectors) and local ancestry;<sup>1-6</sup> 17 SNPs were directly genotyped, while 2 were imputed using MACH ( $r^2 > 0.98$ ). All 19 variants were common ( $\geq 0.05$ ) in African Americans, with 11 variants being more common in Europeans than African Americans (Figure 1). In previous GWAS, the index signals had very modest odds ratios (1.05-1.29 per copy of the risk allele) and our sample size provided  $\geq 70\%$  statistical power to detect the reported effects for 12 of the 19 variants (at  $P < 0.05$ ). We observed positive associations with 11 of the 19 variants (OR  $> 1$ ) however only 4 were statistically significant ( $P < 0.05$  at 2q35, 9q31, 10q26 and 19p13). Of the 15 variants that were not replicated at  $P < 0.05$ , statistical power was  $< 70\%$  for only 7 of the variants. Although power was more limited, we also evaluated associations by estrogen receptor (ER) status as some risk variants have been found to be more strongly associated with ER-positive (ER+) or ER-negative (ER-) breast cancer. We observed positive associations with 12 variants (2 at  $P < 0.05$ ) for ER+ disease ( $n=1,520$ ) and with 9 variants for ER- (3 at  $P < 0.05$ ;  $n=988$ ). For only one variant did we observe statistically significant risk heterogeneity by ER status (rs13387042 at 2q35,  $P=0.013$ ).

Aside from statistical power, the lack of a statistically significant association with an index variant (OR  $> 1$  and  $p < 0.05$ ) suggests that the particular variant revealed in the GWAS populations may not be adequately correlated with the biologically relevant allele in African Americans. In an attempt to identify a better genetic marker of risk in African Americans we conducted fine-mapping across all risk regions using genotyped SNPs on the Illumina 1M array and imputed SNPs to Phase 2 HapMap populations. If a marker associated with risk in African Americans represents the same signal as that reported in the initial GWAS, then it should be correlated to some degree with the index signal in the GWAS population. Using HapMap data

for the populations in which the risk variant was identified (Utah residents with ancestry from northern and western Europe (CEU), or Han Chinese in Beijing, China (CHB)), we catalogued and tested all SNPs that were correlated ( $r^2 \geq 0.2$ ) with the index signal (within 250kb), applying an  $\alpha_a$  of  $3.2 \times 10^{-3}$  which was estimated as 0.05/the average number of tags needed to capture ( $r^2 \geq 0.8$ ) the common risk alleles correlated with the index allele in each region in the Yoruba HapMap population (in Ibadan, Nigeria (YRI)). We also tested for novel independent associations, focusing on SNPs that were uncorrelated with the index signal in the initial GWAS populations. Here, we applied a Bonferroni correction for defining novel associations as statistically significant in each region, with  $\alpha_b$  estimated as 0.05/the total number of tags needed to capture ( $r^2 \geq 0.8$ ) all common risk alleles in the 19 regions in the YRI population ( $\alpha_b = 1.0 \times 10^{-5}$ ; similar to the genome-wide-type correction of  $5 \times 10^{-8}$ , which accounts for the number of tags needed to capture all common alleles in the genome). For each region, stepwise logistic regression was used with SNPs kept in the final model based on  $\alpha_a$  or  $\alpha_b$ . These procedures were applied to all cases and controls as well as in hypothesis-generating analyses stratified by ER status. At 9 loci we detected variants that were statistically significantly associated with breast cancer risk in African Americans. These regions include 9q31 where the sole marker of risk was the index signal (rs865686: OR=1.08; P=0.034). Through fine-mapping we revealed markers in four regions that were more significantly associated with risk than the index signal (>1 order of magnitude change in the p-value) and are likely capturing the same signal (2q35, 5q11, 10q26 and 19p13). We also identified markers in four regions that are not correlated with the index signal in the GWAS populations (8q24, 10q22, 11q13 and 16q12) and may represent putative novel risk variants, with one being specific for ER+ disease (8q24). These regions are discussed below.

### ***Risk variants that better define the index signal in African Americans***

#### ***2q35***

The index signal at 2q35 was statistically significantly associated with risk of overall breast cancer (rs13387042: OR=1.12, P= $7.5 \times 10^{-3}$ ) and ER+ disease (OR=1.22, P= $2.6 \times 10^{-4}$ ). However, we found stronger associations with two markers that are each modestly correlated with the index signal in CEU and YRI: rs13000023 with overall breast cancer (OR=1.20, P= $5.8 \times 10^{-4}$ ) and rs12998806: with ER+ disease (OR=1.39, P= $3.3 \times 10^{-6}$ ). The signal in this region appeared limited to ER+ breast cancer, which is consistent with the initial report of this risk locus.<sup>3</sup>

#### ***5q11***

We found a positive non-significant association with the index signal at 5q11, which is located 79 kb centromeric of the *MAP3K1* gene (rs889312: OR=1.07, P=0.084). Fine-mapping revealed statistically significant associations with markers, rs16886165 for overall breast cancer (OR=1.15, P= $6.5 \times 10^{-4}$ ) and rs832529 for ER- disease (OR=1.22, P= $1.3 \times 10^{-3}$ ). These SNPs show greater correlation with the index signal in Europeans (CEU,  $r^2=0.40$  and  $0.46$ ) than in Africans (YRI,  $r^2 < 0.01$  and  $r^2=0.09$ ), which suggests that they may be better markers of the biologically functional variant in African Americans.

#### ***10q26***

Both the index signal, rs2981582 (OR=1.11, P= $8.6 \times 10^{-3}$ ), and rs2981578, that was identified previously through fine-mapping in African Americans (which some of these studies contributed to)<sup>7</sup>, were statistically significantly associated with risk (OR=1.24, P= $1.7 \times 10^{-4}$ ). Variant rs2981578 was the most strongly associated marker in the region for overall breast cancer and for ER+ disease, which is consistent with previous reports of variation in this region being more strongly associated with ER+ breast cancer.<sup>8</sup> In fine-mapping the locus we observed a suggestive association with a correlated marker and ER- disease (rs2912774: OR=1.19, P= $2.1 \times 10^{-3}$ ) however the association was also noted with ER+ disease (OR=1.10, P=0.041) and is likely capturing the same signal as rs2981578.

### 19p13

19p13 was the first risk locus reported to harbor a variant that may be specific for ER- disease.<sup>9</sup> In African Americans, the index variant was statistically significantly associated with risk of overall breast cancer (rs2363956: OR=1.14,  $P=8.0\times 10^{-4}$ ), as well as ER+ (OR=1.12,  $P=0.016$ ) and ER- disease (OR=1.14,  $P=0.01$ ). The most significant association in the region for overall breast cancer and ER+ disease was with rs3745185 ( $P=3.7\times 10^{-5}$  and  $P=8.2\times 10^{-4}$ , respectively), which is likely to be capturing the same functional variant ( $r^2=0.57$  in CEU and 0.19 in YRI). The most significant marker for ER- breast cancer was correlated with both rs2363956 and rs3745185 (rs11668840: OR=1.25,  $P=5.1\times 10^{-5}$ ).

### ***Novel risk-associated markers at breast cancer susceptibility loci.***

#### 8q24

Given the importance of the 8q24 locus in cancer, we conducted association testing across the entire cancer risk region (126.0 Mb-130.0 Mb).<sup>10,11</sup> The index signal (rs13281615) was not statistically significantly associated with risk in African Americans, nor did we identify significant associations with correlated SNPs. However, we did detect a significant association with rs16902056 and ER+ breast cancer (risk allele frequency, 0.95;  $P=6.7\times 10^{-6}$ ; ER-:  $P=0.66$ ). This SNP is located 78 kb centromeric of the index variant and is not correlated with the index variant ( $r^2<0.01$  in CEU and  $r^2=0.027$  in YRI). No statistically significant associations were observed with variants found previously in association with cancers of the bladder and ovary, or leukemia (rs9642880: OR=1.03,  $P=0.58$ ; rs10088218: OR=1.02,  $P=0.62$ ; rs2456449: OR=1.07,  $P=0.14$ ). Of the known risk variants for prostate cancer we found a single nominally significant ( $P<0.05$ ) association with the same risk allele of rs1016343 ( $P=0.015$ ) which is located >260 kb centromeric of the breast cancer risk region and is not correlated with rs13281615 or rs16902056.

#### 10q22

We observed no association with the index signal at 10q22 (rs704010) which is located in intron 1 of the gene *ZMIZ1*, or with any correlated markers. However, we did detect strong evidence of a second signal located 215 kb telomeric in intron 12 of the gene *ZMIZ1* (rs12355688: OR=1.24,  $P=6.8\times 10^{-6}$ ). This putative novel risk variant is not correlated with the index variant in the CEU or YRI populations ( $r^2<0.01$ ).

#### 11q13

No positive association was noted with the index variant at 11q13. However, we did detect evidence of a second independent signal (rs609275: OR=1.20,  $P=1.0\times 10^{-5}$ ), located 74 kb telomeric, and 53 kb centromeric of *CCND1*. The variant is monomorphic and uncorrelated with the index signal in the CEU population; and  $r^2$  with the index signal in the YRI population is  $<0.01$ .

#### 16q12

As in previous studies of African Americans we were not able to replicate the association signal defined by the index variant rs3803662.<sup>12,13</sup> A recent study of African Americans reported a suggestive association with SNP rs3104746, which is located 15 kb telomeric of rs3803662.<sup>14</sup> This SNP has a minor allele frequency of 0.04 in the HapMap CEU population, 0.19 in our African American controls, and is modestly correlated with rs3803662 in Africans ( $r^2=0.31$  in YRI), but not in Europeans ( $r^2=0.038$ ). Fine-mapping around this putative signal revealed a perfect proxy ( $r^2=1$ ) for rs3104746, rs3112572, which is significantly associated with breast cancer risk in African Americans (OR=1.18,  $P=3.9\times 10^{-4}$ ) with the association noted to be stronger for ER+ breast cancer (OR=1.27,  $P=3.1\times 10^{-5}$ ).

For index SNPs found to be nominally associated with breast cancer risk, as well as risk-associated markers identified through fine-mapping, we also tested for associations by genotype. Results from the genotype-specific model were consistent with log-additive-

associations. Risk variants at 2q35 and 8q24 were also found to have significantly stronger associations with ER+ breast cancer than ER- disease which is consistent with previous studies.<sup>8</sup>

We observed no statistically significant associations with common variation at 10 risk loci on 1p11, 3p24, 5p12, 6q25, 9p21, 10p15, 10q21, 11p15, 14q24 and 17q23.

### ***Risk modeling***

We next estimated the cumulative effect of all breast cancer risk variants, and compared a summary risk score comprised of unweighted counts of all GWAS reported risk variants to a risk score that included variants we identified as being associated with risk in African Americans. Using the 19 index signals from GWAS, the risk per allele was 1.04 (95% CI, 1.02-1.06;  $P=6.1\times 10^{-5}$ ) and individuals in the top quintile of the risk allele distribution were at 1.4-fold greater risk ( $P=7.4\times 10^{-5}$ ) of breast cancer compared to those in the lowest quartile. As expected, the risk score was improved when utilizing the markers that we identified at the known risk loci as being more relevant to African Americans (8 alleles for overall breast cancer: 2q35, 5q11, 9q31, 10q22, 10q26, 11q13, 16q12 and 19p13; OR=1.18; 95% CI, 1.14-1.22;  $P=2.8\times 10^{-24}$ ), with risk for those in the top quartile being 2.2-times that observed in the lowest quintile ( $P=3.6\times 10^{-17}$ ). We observed an increase of 1.9 percentage points in the area under the curve (AUC) ( $P=2.6\times 10^{-6}$ ). This score was significantly associated with risk of both ER+ (OR=1.20,  $P=1.7\times 10^{-19}$ ) and ER- (OR=1.15,  $P=2.8\times 10^{-9}$ ) disease ( $P_{\text{het}}=0.12$ ).

Stratifying by first-degree family history of breast cancer differentiated risk further with those with a family history and in the top quintile of the risk score distribution (4% of the population) having a 3.4-fold greater risk ( $P=9.9\times 10^{-14}$ ) compared to those without a family history and in the lowest quintile of the risk score.

In hypothesis-generating analyses, we also developed risk scores for ER+ and ER- breast tumor subtypes utilizing the most informative markers revealed through fine-mapping of each phenotype. These phenotype-specific scores were highly significant (ER+: OR=1.30,  $P=6.0\times 10^{-18}$ ; ER-: OR=1.20,  $P=2.3\times 10^{-10}$ ) with statistically significant heterogeneity noted when the scores were applied to the other subtype ( $P_{\text{het}}=1.7\times 10^{-5}$  and  $5.0\times 10^{-3}$  for ER+ and ER- scores, respectively).

### ***Summary***

In this large study of breast cancer in African American women we were able to replicate associations with 4 of the 19 index variants (at  $P<0.05$ ). Through fine-mapping, we observed that overall breast cancer risk was statistically significantly associated with markers in 4 regions which are likely to capture the GWAS-reported signal and to serve as better markers of the functional allele and risk in African Americans. We also detected putative novel associations that are independent of the index signals in 3 regions for overall breast cancer (10q22, 11q13 and 16q12) and in one region for ER+ disease (8q24). In 10 of the risk regions, however, we were not able to replicate the GWAS index signals, nor did we detect statistically significant associations of common SNPs with breast cancer risk at the levels of statistical significance we set for fine-mapping.

In the four regions we observed risk markers that are correlated and in the same LD block with the index markers in CEU (rs13000023 at 2q35, rs16886165 at 5q11, rs2981578 at 10q26 and rs3745185 at 19p13;  $r^2\geq 0.35$ ). It is likely that these risk markers capture the same signal as the index markers do, however we cannot rule out the possibility that some of them may represent a second, independent signal in the same region. Since the  $r^2$  between these markers and the index markers are higher, the second signals they pick up are likely to be in the same LD as the index signal in CEU, which would be captured by the index marker in the original GWAS, which justified the more liberal P-value we used to select them.



In the four regions where we observed independent signals, the risk alleles (rs16902056 at 8q24, rs12355688 at 10q22, rs609275 at 11q13 and rs3112572 at 16q12) were uncorrelated with ( $r^2 < 0.04$ ) and not in the same LD block as the index variant in Europeans (CEU) [distances from the index signal ranged from 14kb at 16q12 to 215kb at 10q22]. Therefore, these variants are likely to pick up a novel signal independent of the index signal. At 10q22, both the index SNP and the novel variant are located within introns of the *ZMIZ1* gene. *ZMIZ1* encodes zinc finger MIZ-type containing 1, which regulates the activity of various transcription factors including the androgen receptor, Smad3/4, and p53. At 11q13, rs609275 lies 74 kb telomeric of the index signal and in closer proximity to a number of candidate genes including *CCND1* (encoding cyclin D1), a protein crucial for cell cycle control, *ORAOV1* (encoding oral cancer overexpressed 1) and *FGF19* (encoding fibroblast growth factor 19). The association at 16q12 confirms the findings of a previous, smaller study of African Americans,<sup>15</sup> and is consistent with a previous fine-mapping study suggesting that African Americans may harbor a separate causal variant in this region.<sup>12</sup> Whether this variant is influencing the same genes/pathways as the index variant rs3803662 is not known, however the stronger associations noted for both variants with ER+ disease suggest that they may affect the same biological process.

Notably, at region 19p13 which was originally reported in association with ER- breast cancer,<sup>9</sup> the index signal was statistically significantly associated with both ER+ and ER- subtypes in African Americans. In addition, we found a stronger marker in this region (rs3745185) for ER+ as well as overall breast cancer risk. We also found stronger associations with ER+ than ER- disease for variants in many regions, including 2q35, 8q24, 10q26 and 16q12, which is consistent with previous reports.<sup>8</sup> We also found strong signals for ER- disease in regions 5q11, 10q26 and 19p13. While there haven't been any reported associations between these signals and ER- disease in European-ancestry populations, it is possible that they explain some of the excess risk for ER- disease in African-Americans, since these risk alleles have higher frequencies in this population than they do in European-ancestry populations.

The majority of the variants identified by GWAS for common cancers are of low risk (relative risks  $< 1.30$ ) and in aggregate are not yet informative for risk prediction. Until the functional alleles at each susceptibility locus are identified and their effects are accurately estimated, modeling of the genetic risk will rely on markers that best capture risk for a given population. Many of the markers we identified at these risk loci appear to provide improvement over the GWAS-identified variants in defining African American women who are at greater risk of breast cancer. The risk score for overall breast cancer was also equally efficient for ER+ and ER- tumors. However, our hypothesis-generating model suggests that identification of tumor subtype-specific variants will improve the fit of these models.

While this is the largest study of African Americans to date to investigate genetic risk at known breast cancer susceptibility loci, statistical power was still limited. We had only 35% power to detect an OR of 1.10 for a risk allele of 0.10 frequency, which may account for our inability to replicate GWAS signals or risk-associated markers in 10 of the regions. While attempting to apply a strict threshold for declaring significance through fine-mapping, we did not take into account testing for multiple phenotypes (overall breast as well as ER+ and ER- disease). As a result, the  $\alpha$  levels used as selection criteria may be too liberal. However, our risk modeling focused on the variants revealed for overall breast cancer, whereas we consider the associations observed for markers identified for ER+ or ER- disease and used in the subtype-specific risk modeling as hypothesis-generating. Since all of the cases and controls used for fine-mapping/discovery were also included in the risk modeling, the risk model is likely to overestimate the level of prediction due to winner's curse. Instead of partitioning the sample into test and validation sets, we felt it was necessary to use all of the subjects in the association testing of known variants and in fine-mapping to increase the statistical power to detect associations in each region. Therefore, other comparably large studies of African Americans must be performed in the future to test the model presented.

### **Meta-Analyses with other Breast Cancer GWAS**

We have also initiated meta-analyses with other GWAS including a GWAS of triple negative breast cancer (PI, Fergus Couch) and a GWAS of estrogen receptor negative breast from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). From these meta-analyses we have revealed a novel risk locus for estrogen receptor negative breast cancer.

### ***A Novel Risk Locus for Estrogen Receptor Negative Breast Cancer***

Compared to women of European ancestry, women of African descent are more likely to be diagnosed with estrogen receptor negative breast cancer.<sup>16</sup> ER negative and triple negative tumors, which are deficient in the expression of estrogen, progesterone (PR) and human epidermal growth factor-2 (HER2) receptors, are observed at even higher rates among African women in Africa<sup>17</sup>, suggesting a genetic component to the high risk of ER negative phenotypes in women of African descent. Similarly, ER negative breast cancers and triple-negative breast cancers are also the predominant histological subtypes in women with germline mutations in BRCA1.<sup>18</sup> The enrichment for ER negative disease in this genetically predisposed population also suggests the existence of additional genetic factors that contribute to the risk of ER negative disease. Support for the presence of these factors was recently provided by a genome-wide association study (GWAS) of breast cancer in BRCA1 mutation carriers, in which a common risk variant for ER negative breast cancer on chromosome 19p13 was identified that also displayed significance in ER negative and triple negative disease in the general population.<sup>9</sup>

To search for genetic risk factors for ER negative breast cancer phenotypes, we combined results from our GWAS of breast cancer in African American women [AABC: 3,016 cases (988 with ER negative disease) and 2,745 controls] with results from a GWAS of triple negative breast cancer in women of European ancestry (TNBCC: 1,718 cases and 3,670 controls). In TNBCC, cases were genotyped with the Illumina 660W array. Genotypes of TNBCC cases were compared with GWAS data for publicly available controls. Both studies imputed genotypes for common SNPs in Phase 2 HapMap populations (release 21). A total of 3,154,485 SNPs, genotyped and imputed were analyzed in stage 1 of the meta-analysis. In the combined results, only SNP 241 at chromosome 5 displayed a genome-wide significant association with ER negative breast cancer (AABC: OR per allele=1.32,  $p=1.9 \times 10^{-6}$ ; TNBCC: OR=1.25,  $p=1.2 \times 10^{-3}$ ; combined OR = 1.29,  $p=1.0 \times 10^{-8}$ ).

To further confirm the association on chromosome 5, we genotyped SNP 241 in women of European ancestry, which included 8,313 cases (1,308 ER negatives) and 10,879 controls from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3) and 6,307 cases (813 ER negatives) and 6,722 controls from Studies of Epidemiology and Risk Factors in Cancer Heredity (SEARCH). Evidence for replication was observed for 241 and ER negative breast cancer in both studies (BPC3: OR=1.10,  $p=0.053$ ; SEARCH: OR=1.21,  $p=1.7 \times 10^{-3}$ ).

In combining the results across all studies (5,874 ER negative cases and 21,389 controls with genotype data), SNP 241 was significantly associated with an increased risk of ER negative breast cancer (OR = 1.19, 95% CI, 1.13-1.25;  $p=1.6 \times 10^{-10}$ ). The risk for heterozygote and homozygote carriers was 1.14 (95 % CI, 1.06-1.23) and 1.46 (95% CI, 1.29-1.65), respectively. We observed little evidence of heterogeneity for the reported association for this variant by study/country in AABC ( $p_{\text{het}}=0.97$ ), TNBCC ( $p_{\text{het}}=0.88$ ) or BPC3 ( $p_{\text{het}}=0.41$ ).

In an analysis of ER positive cases, SNP 241 was only weakly associated with risk in African Americans (AABC: 1,518 ER positive cases and 2,743 controls with genotype data: OR=1.08;  $p=0.10$ ) and in women of European ancestry (BPC3: 4,671 ER positive cases and 10,397 controls, OR=1.03,  $p=0.26$ ; SEARCH: 3,359 ER positive cases and 6,722 controls,

OR=1.02,  $p=0.52$ ; combined for all populations: OR=1.04,  $p=0.06$ ,  $p_{\text{Het}} = 0.61$ ). This result suggests that the association with breast cancer might be specific for ER negative disease ( $P$ -value for case-only test of ER negative versus ER positive =  $4.0 \times 10^{-4}$ ).

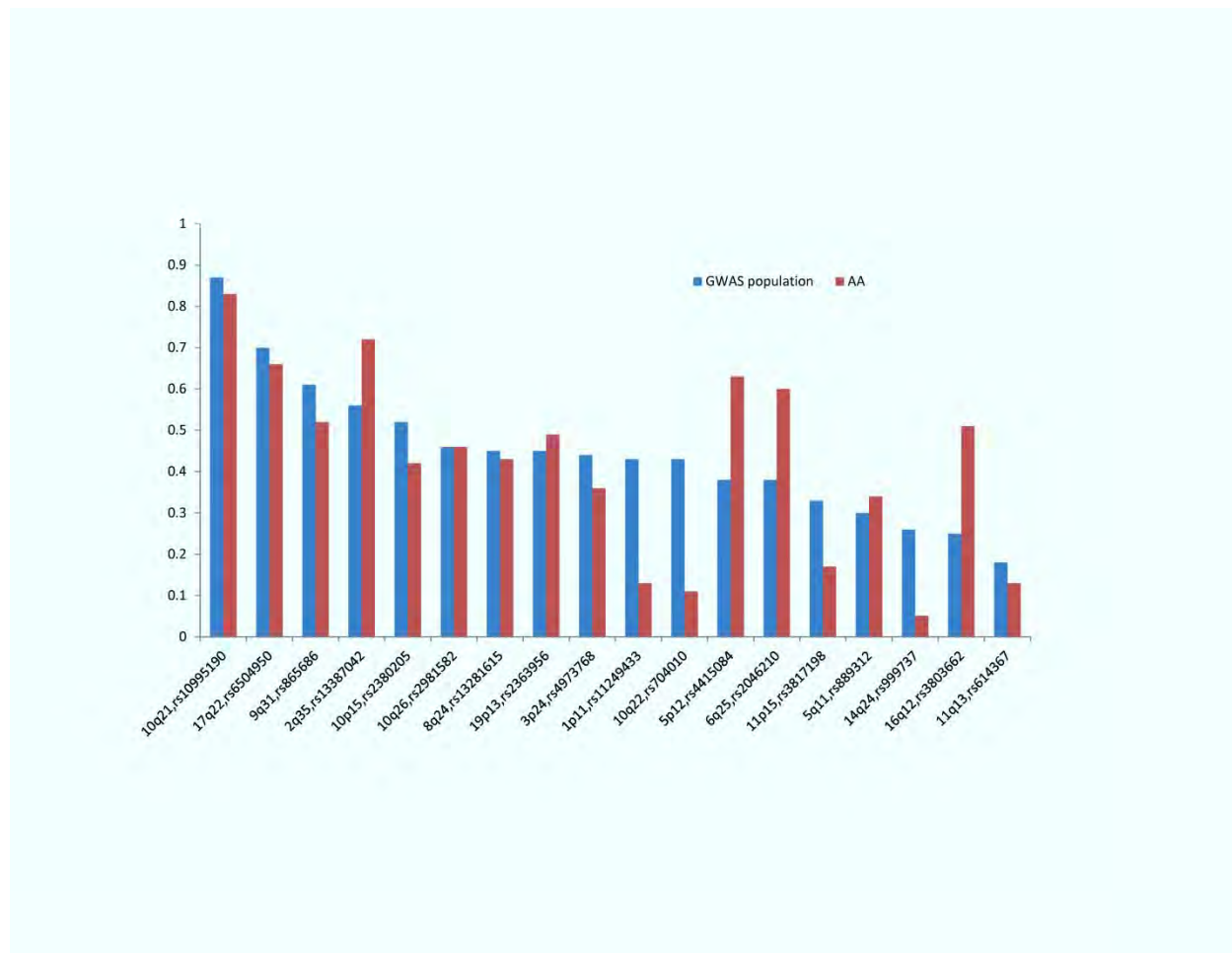
### **Summary**

Similar to 8q24<sup>10,19,20</sup> and 11q13<sup>21-23</sup>, the TERT/CLPTM1L locus harbors multiple risk variants for different cancers (reviewed in <sup>24</sup>). SNP 241 is modestly correlated ( $r^2 = 0.13-0.43$  in 1000 Genomes Project populations of European and African ancestry) with variants found for serous ovarian cancer (rs7726159), glioma (rs2736100), and lung cancer (rs2736100, rs2735940).<sup>25-27</sup> Aside from risk variant rs2853676 found for glioma<sup>27</sup> that was associated with risk in TNBCC ( $p=0.014$ ,  $r^2=0.05$  with SNP 241), none of the known risk variants identified for other cancers in the TERT/CLPTM1L region were significantly associated with breast cancer risk in TNBCC or AABC. The TERT gene encodes the catalytic subunit of telomerase which controls telomere length, a process linked with genomic instability and implicated in tumorigenesis. The TERT locus may highlight another biological process common to the pathogenesis of ER negative breast cancer and serous ovarian cancer that is also shared with other cancers.

A manuscript describing these findings is currently under review at *Nature Genetics*.

**Figure 1. Risk allele frequencies in Europeans and African Americans.**

The distribution of risk allele frequencies (RAF) for the 19 index SNPs in HapMap CEU (CHB for rs2046210) and African Americans (AA). The variants are sorted based on the RAF in CEU.



## Key Research Accomplishments

- 1) We conducted detailed fine-mapping of the known breast cancer risk loci and have validated and improved upon markers of risk that better characterize their contribution to breast cancer risk in women of African ancestry.
- 2) We conducted the largest study to directly investigate genetic susceptibility to ER- breast cancer, which is an aggressive type of breast cancer that is associated with poor prognosis and decreased survival, and for which treatment options are limited.
- 3) Of the 20 risk loci for breast cancer, only the association at chromosome 19 is limited to ER- breast cancer. The chromosome 5 locus becomes only the 2<sup>nd</sup> risk locus that has been identified for ER- breast cancer.
- 4) We have limited knowledge of the etiology of the ER- breast cancer. The association of a common variant at chromosome 5 with ER- disease substantiates epidemiological studies in demonstrating distinct etiologies for breast cancer subtypes as defined by tumor markers. An understanding of the biological pathways is the first step towards developing effective strategies for preventing and treating ER- disease.
- 5) This is the first GWAS of breast cancer in African American women, a population in which the incidence of ER- disease is substantially greater than that in other populations. Our findings suggest that this the chromosome 5 locus may contribute to the greater incidence of ER- breast cancer in this population.

### **Reportable Outcomes**

- The findings presented in this progress report will be discussed at the DOD BCRP meeting in Orlando (August 2011).
- Two papers under review at *Human Molecular Genetics* and *Nature Genetics*.

## Conclusion

Through fine-mapping of the breast cancer susceptibility regions in a large sample of African American women, we identified markers that improve breast cancer risk prediction for this population. In aggregate, the informative markers at the established risk loci allow for an improvement in modeling of breast cancer risk over GWAS-reported markers in African Americans (per allele OR=1.18,  $P=2.8 \times 10^{-24}$  vs. OR=1.04,  $P=6.1 \times 10^{-5}$ ). Validation and enhancement of this model is needed before risk modeling based on genetic variants of low risk can be implemented in the clinical setting. At chromosome 5, the identification of the variant directly responsible for the association will be required to fully address the extent to which this locus contributes to the greater incidence of ER negative and triple negative tumors in women of African ancestry. However, it is notable that both the risk allele frequency and the odds ratio for SNP 241 are greater in African American women (frequency, 0.57; OR=1.32,  $p=1.9 \times 10^{-6}$ ) than in women of European ancestry (frequency, 0.26; OR=1.15,  $p=2.2 \times 10^{-6}$ ). Based on these differences in frequency and effect size, and assuming this variant is an equally good surrogate for the biologically functional allele in each population, we estimate that this locus may be responsible for a notable fraction (25%, 95% CI, 7-49%) of the greater incidence rate of ER negative breast cancer in women of African than European ancestry. Larger studies with well-characterized tumor pathology information will be needed to determine if the association we observed applies to all ER negative disease or tumor subtypes that include ER negative status as a component, such as triple negative breast cancer. Our findings provide further support for the presence of genetic susceptibility to ER negative breast cancer and demonstrate the importance of discovery efforts in multiple populations.

A future direction of our work will be to continue to combine GWAS data from multiple populations (as we have done for the chromosome 5 locus). Such work is in progress and in preliminary analyses we have found 2 additional novel loci on chromosomes 6 and 20 which we are currently pursuing in multiethnic replication studies.

Revealing the genetic causes of breast cancer *in each population* will in time translate into more targeted preventive measures and treatment strategies for those at risk of developing the breast cancer. The risk locus on chromosome 5 is this first clue that there may be a genetic basis for the greater incidence of estrogen receptor negative breast cancer in women of African ancestry.

## References

1. Easton DF, P.K., Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
2. Hunter DJ, K.P., Jacobs KB, Cox DG, Yeager M, Hankinson, SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-4 (2007).
3. Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., Aben, K.K., Strobbe, L.J., Albers-Akkers, M.T., Swinkels, D.W., Henderson, B.E., Kolonel, L.N., Le Marchand, L., Millastre, E., Andres, R., Godino, J., Garcia-Prats, M.D., Polo, E., Tres, A., Mouy, M., Saemundsdottir, J., Backman, V.M., Gudmundsson, L., Kristjansson, K., Bergthorsson, J.T., Kostic, J., Frigge, M.L., Geller, F., Gudbjartsson, D., Sigurdsson, H., Jonsdottir, T., Hrafnkelsson, J., Johannsson, J., Sveinsson, T., Myrdal, G., Grimsson, H.N., Jonsson, T., von Holst, S., Werelius, B., Margolin, S., Lindblom, A., Mayordomo, J.I., Haiman, C.A., Kiemeny, L.A., Johannsson, O.T., Gulcher, J.R., Thorsteinsdottir, U., Kong, A. & Stefansson, K. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-9 (2007).
4. Stacey, S.N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S.A., Jonsson, G.F., Jakobsdottir, M., Bergthorsson, J.T., Gudmundsson, J., Aben, K.K., Strobbe, L.J., Swinkels, D.W., van Engelenburg, K.C., Henderson, B.E., Kolonel, L.N., Le Marchand, L., Millastre, E., Andres, R., Saez, B., Lambea, J., Godino, J., Polo, E., Tres, A., Picelli, S., Rantala, J., Margolin, S., Jonsson, T., Sigurdsson, H., Jonsdottir, T., Hrafnkelsson, J., Johannsson, J., Sveinsson, T., Myrdal, G., Grimsson, H.N., Sveinsdottir, S.G., Alexiusdottir, K., Saemundsdottir, J., Sigurdsson, A., Kostic, J., Gudmundsson, L., Kristjansson, K., Masson, G., Fackenthal, J.D., Adebamowo, C., Ogundiran, T., Olopade, O.I., Haiman, C.A., Lindblom, A., Mayordomo, J.I., Kiemeny, L.A., Gulcher, J.R., Rafnar, T., Thorsteinsdottir, U., Johannsson, O.T., Kong, A. & Stefansson, K. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **40**, 703-6 (2008).
5. Thomas, G., Jacobs, K.B., Kraft, P., Yeager, M., Wacholder, S., Cox, D.G., Hankinson, S.E., Hutchinson, A., Wang, Z., Yu, K., Chatterjee, N., Garcia-Closas, M., Gonzalez-Bosquet, J., Prokunina-Olsson, L., Orr, N., Willett, W.C., Colditz, G.A., Ziegler, R.G., Berg, C.D., Buys, S.S., McCarty, C.A., Feigelson, H.S., Calle, E.E., Thun, M.J., Diver, R., Prentice, R., Jackson, R., Kooperberg, C., Chlebowski, R., Lissowska, J., Peplonska, B., Brinton, L.A., Sigurdson, A., Doody, M., Bhatti, P., Alexander, B.H., Buring, J., Lee, I.M., Vatten, L.J., Hveem, K., Kumle, M., Hayes, R.B., Tucker, M., Gerhard, D.S., Fraumeni, J.F., Jr., Hoover, R.N., Chanock, S.J. & Hunter, D.J. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**, 579-84 (2009).
6. Zheng, W., Long, J., Gao, Y.T., Li, C., Zheng, Y., Xiang, Y.B., Wen, W., Levy, S., Deming, S.L., Haines, J.L., Gu, K., Fair, A.M., Cai, Q., Lu, W. & Shu, X.O. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-8 (2009).



7. Udler, M.S., Meyer, K.B., Pooley, K.A., Karlins, E., Struewing, J.P., Zhang, J., Doody, D.R., MacArthur, S., Tyrer, J., Pharoah, P.D., Luben, R., Bernstein, L., Kolonel, L.N., Henderson, B.E., Le Marchand, L., Ursin, G., Press, M.F., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.Y., Wu, P.E., Wang, H.C., Kang, D., Yoo, K.Y., Noh, D.Y., Ahn, S.H., Ponder, B.A., Haiman, C.A., Malone, K.E., Dunning, A.M., Ostrander, E.A. & Easton, D.F. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum Mol Genet* **18**, 1692-703 (2009).
8. Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., Richesson, D.A., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Arias, J.I., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Zamora, P., Brauch, H., Justenhoven, C., Hamann, U., Ko, Y.D., Bruening, T., Haas, S., Dork, T., Schurmann, P., Hillemanns, P., Bogdanova, N., Bremer, M., Karstens, J.H., Fagerholm, R., Aaltonen, K., Aittomaki, K., von Smitten, K., Blomqvist, C., Mannermaa, A., Uusitupa, M., Eskelinen, M., Tengstrom, M., Kosma, V.M., Kataja, V., Chenevix-Trench, G., Spurdle, A.B., Beesley, J., Chen, X., Devilee, P., van Asperen, C.J., Jacobi, C.E., Tollenaar, R.A., Huijts, P.E., Klijn, J.G., Chang-Claude, J., Kropp, S., Slinger, T., Flesch-Janys, D., Mutschelknauss, E., Salazar, R., Wang-Gohrke, S., Couch, F., Goode, E.L., Olson, J.E., Vachon, C., Fredericksen, Z.S., Giles, G.G., Baglietto, L., Severi, G., Hopper, J.L., English, D.R., Southey, M.C., Haiman, C.A., Henderson, B.E., Kolonel, L.N., Le Marchand, L., Stram, D.O., Hunter, D.J., Hankinson, S.E., Cox, D.G., Tamimi, R., Kraft, P., Sherman, M.E., Chanock, S.J., Lissowska, J., Brinton, L.A., Peplonska, B., Klijn, J.G., Hooning, M.J., Meijers-Heijboer, H., Collee, J.M., van den Ouweland, A., Uitterlinden, A.G., Liu, J., Lin, L.Y., Yuqing, L., Humphreys, K., Czene, K., Cox, A., Balasubramanian, S.P., Cross, S.S., Reed, M.W., Blows, F., Driver, K., Dunning, A., Tyrer, J., Ponder, B.A., Sangrajrang, S., Brennan, P., McKay, J., Odefrey, F., Gaborieau, V., Sigurdson, A., Doody, M., Struewing, J.P., Alexander, B., Easton, D.F. & Pharoah, P.D. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* **4**, e1000054 (2008).
9. Antoniou, A.C., Wang, X., Fredericksen, Z.S., McGuffog, L., Tarrell, R., Sinilnikova, O.M., Healey, S., Morrison, J., Kartsonaki, C., Lesnick, T., Ghousaini, M., Barrowdale, D., Peock, S., Cook, M., Oliver, C., Frost, D., Eccles, D., Evans, D.G., Eeles, R., Izatt, L., Chu, C., Douglas, F., Paterson, J., Stoppa-Lyonnet, D., Houdayer, C., Mazoyer, S., Giraud, S., Lasset, C., Remenieras, A., Caron, O., Hardouin, A., Berthet, P., Hogervorst, F.B., Rookus, M.A., Jager, A., van den Ouweland, A., Hoogerbrugge, N., van der Luit, R.B., Meijers-Heijboer, H., Gomez Garcia, E.B., Devilee, P., Vreeswijk, M.P., Lubinski, J., Jakubowska, A., Gronwald, J., Huzarski, T., Byrski, T., Gorski, B., Cybulski, C., Spurdle, A.B., Holland, H., Goldgar, D.E., John, E.M., Hopper, J.L., Southey, M., Buys, S.S., Daly, M.B., Terry, M.B., Schmutzler, R.K., Wappenschmidt, B., Engel, C., Meindl, A., Preisler-Adams, S., Arnold, N., Niederacher, D., Sutter, C., Domchek, S.M., Nathanson, K.L., Rebbeck, T., Blum, J.L., Piedmonte, M., Rodriguez, G.C., Wakeley, K., Boggess, J.F., Basil, J., Blank, S.V., Friedman, E., Kaufman, B., Laitman, Y., Milgrom, R., Andrulis, I.L., Glendon, G., Ozcelik, H., Kirchoff, T., Vijai, J., Gaudet, M.M., Altshuler, D., Guiducci, C., Loman, N., Harbst, K., Rantala, J., Ehrencrona, H., Gerdes, A.M., Thomassen, M., Sunde, L., Peterlongo, P., Manoukian, S., Bonanni, B., Viel, A., Radice, P., Caldes, T., de la Hoya, M., Singer, C.F., Fink-Retter, A., Greene, M.H., Mai, P.L., Loud, J.T.,

- Guidugli, L., Lindor, N.M., Hansen, T.V., Nielsen, F.C., Blanco, I., Lazaro, C., Garber, J., Ramus, S.J., Gayther, S.A., Phelan, C., Narod, S., Szabo, C.I., Benitez, J., Osorio, A., Nevanlinna, H., Heikkinen, T., Caligo, M.A., Beattie, M.S., Hamann, U., Godwin, A.K., Montagna, M., Casella, C., Neuhausen, S.L., Karlan, B.Y., Tung, N., Toland, A.E., Weitzel, J., Olopade, O., Simard, J., Soucy, P., Rubinstein, W.S., Arason, A., Rennert, G., Martin, N.G., Montgomery, G.W., Chang-Claude, J., Flesch-Janys, D., Brauch, H., Severi, G., Baglietto, L., Cox, A., Cross, S.S., Miron, P., Gerty, S.M., Tapper, W., Yannoukakos, D., Fountzilas, G., Fasching, P.A., Beckmann, M.W., Dos Santos Silva, I., Peto, J., Lambrechts, D., Paridaens, R., Rudiger, T., Forsti, A., Winqvist, R., Pylkas, K., Diasio, R.B., Lee, A.M., Eckel-Passow, J., Vachon, C., Blows, F., Driver, K., Dunning, A., Pharoah, P.P., Offit, K., Pankratz, V.S., Hakonarson, H., Chenevix-Trench, G., Easton, D.F. & Couch, F.J. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* **42**, 885-92 (2010).
10. Haiman, C.A., Le Marchand, L., Yamamoto, J., Stram, D.O., Sheng, X., Kolonel, L.N., Wu, A.H., Reich, D. & Henderson, B.E. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* **39**, 954-6 (2007).
  11. Haiman, C.A., Patterson, N., Freedman, M.L., Myers, S.R., Pike, M.C., Waliszewska, A., Neubauer, J., Tandon, A., Schirmer, C., McDonald, G.J., Greenway, S.C., Stram, D.O., Le Marchand, L., Kolonel, L.N., Frasco, M., Wong, D., Pooler, L.C., Ardlie, K., Oakley-Girvan, I., Whittemore, A.S., Cooney, K.A., John, E.M., Ingles, S.A., Altshuler, D., Henderson, B.E. & Reich, D. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* **39**, 638-44 (2007).
  12. Udler, M.S., Ahmed, S., Healey, C.S., Meyer, K., Struewing, J., Maranian, M., Kwon, E.M., Zhang, J., Tyrer, J., Karlins, E., Platte, R., Kalmyrzaev, B., Dicks, E., Field, H., Maia, A.T., Prathalingam, R., Teschendorff, A., McArthur, S., Doody, D.R., Luben, R., Caldas, C., Bernstein, L., Kolonel, L.K., Henderson, B.E., Wu, A.H., Le Marchand, L., Ursin, G., Press, M.F., Lindblom, A., Margolin, S., Shen, C.Y., Yang, S.L., Hsiung, C.N., Kang, D., Yoo, K.Y., Noh, D.Y., Ahn, S.H., Malone, K.E., Haiman, C.A., Pharoah, P.D., Ponder, B.A., Ostrander, E.A., Easton, D.F. & Dunning, A.M. Fine scale mapping of the breast cancer 16q12 locus. *Hum Mol Genet* **19**, 2507-15 (2010).
  13. Zheng, W., Cai, Q., Signorello, L.B., Long, J., Hargreaves, M.K., Deming, S.L., Li, G., Li, C., Cui, Y. & Blot, W.J. Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev* **18**, 2761-4 (2009).
  14. Ruiz-Narvaez, E.A., Rosenberg, L., Cozier, Y.C., Cupples, L.A., Adams-Campbell, L.L. & Palmer, J.R. Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. *Cancer Epidemiology Biomarkers & Prevention* **19**, 1320-7 (2010).
  15. Ruiz-Narvaez, E.A., Rosenberg, L., Cozier, Y.C., Cupples, L.A., Adams-Campbell, L.L. & Palmer, J.R. Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. *Cancer Epidemiol Biomarkers Prev* **19**, 1320-7 (2010).
  16. Carey, L.A., Perou, C.M., Livasy, C.A., Dressler, L.G., Cowan, D., Conway, K., Karaca, G., Troester, M.A., Tse, C.K., Edmiston, S., Deming, S.L., Geradts, J., Cheang, M.C., Nielsen, T.O., Moorman, P.G., Earp, H.S. & Millikan, R.C. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Jama* **295**, 2492-502 (2006).

17. Huo, D., Ikpatt, F., Khramtsov, A., Dangou, J.M., Nanda, R., Dignam, J., Zhang, B., Grushko, T., Zhang, C., Oluwasola, O., Malaka, D., Malami, S., Odetunde, A., Adeoye, A.O., Iyare, F., Falusi, A., Perou, C.M. & Olopade, O.I. Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. *J Clin Oncol* **27**, 4515-21 (2009).
18. Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Eystein Lonning, P. & Borresen-Dale, A.L. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-74 (2001).
19. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struwing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., Meyer, K.B., Haiman, C.A., Kolonel, L.K., Henderson, B.E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.Y., Wu, P.E., Wang, H.C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K.Y., Noh, D.Y., Ahn, S.H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y.L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R.A., Jacobi, C.E., Devilee, P., Klijn, J.G., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G.G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y.D., Spurdle, A.B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V.M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R. & Ponder, B.A. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93 (2007).
20. Kiemeny, L.A., Thorlacius, S., Sulem, P., Geller, F., Aben, K.K., Stacey, S.N., Gudmundsson, J., Jakobsdottir, M., Bergthorsson, J.T., Sigurdsson, A., Blondal, T., Witjes, J.A., Vermeulen, S.H., Hulsbergen-van de Kaa, C.A., Swinkels, D.W., Ploeg, M., Cornel, E.B., Vergunst, H., Thorgeirsson, T.E., Gudbjartsson, D., Gudjonsson, S.A., Thorleifsson, G., Kristinsson, K.T., Mouy, M., Snorraddottir, S., Placidi, D., Campagna, M., Arici, C., Koppova, K., Gurzau, E., Rudnai, P., Kellen, E., Polidoro, S., Guarrera, S., Sacerdote, C., Sanchez, M., Saez, B., Valdivia, G., Ryk, C., de Verdier, P., Lindblom, A., Golka, K., Bishop, D.T., Knowles, M.A., Nikulasson, S., Petursdottir, V., Jonsson, E., Geirsson, G., Kristjansson, B., Mayordomo, J.I., Steineck, G., Porru, S., Buntinx, F., Zeegers, M.P., Fletcher, T., Kumar, R., Matullo, G., Vineis, P., Kiltie, A.E., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., Rafnar, T. & Stefansson, K. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet* **40**, 1307-12 (2008).
21. Purdue, M.P., Johansson, M., Zelenika, D., Toro, J.R., Scelo, G., Moore, L.E., Prokhortchouk, E., Wu, X., Kiemeny, L.A., Gaborieau, V., Jacobs, K.B., Chow, W.H., Zaridze, D., Matveev, V., Lubinski, J., Trubicka, J., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Bucur, A., Bencko, V., Foretova, L., Janout, V., Boffetta, P., Colt, J.S., Davis, F.G., Schwartz, K.L., Banks, R.E., Selby, P.J., Harnden, P., Berg, C.D., Hsing, A.W., Grubb, R.L., 3rd, Boeing, H., Vineis, P., Clavel-Chapelon, F., Palli, D., Tumino, R., Krogh, V., Panico, S., Duell, E.J., Quiros, J.R., Sanchez, M.J., Navarro, C., Ardanaz, E., Dorronsoro,

- M., Khaw, K.T., Allen, N.E., Bueno-de-Mesquita, H.B., Peeters, P.H., Trichopoulos, D., Linseisen, J., Ljungberg, B., Overvad, K., Tjonneland, A., Romieu, I., Riboli, E., Mukeria, A., Shangina, O., Stevens, V.L., Thun, M.J., Diver, W.R., Gapstur, S.M., Pharoah, P.D., Easton, D.F., Albanes, D., Weinstein, S.J., Virtamo, J., Vatten, L., Hveem, K., Njolstad, I., Tell, G.S., Stoltenberg, C., Kumar, R., Koppova, K., Cussenot, O., Benhamou, S., Oosterwijk, E., Vermeulen, S.H., Aben, K.K., van der Marel, S.L., Ye, Y., Wood, C.G., Pu, X., Mazur, A.M., Boulygina, E.S., Chekanov, N.N., Foglio, M., Lechner, D., Gut, I., Heath, S., Blanche, H., Hutchinson, A., Thomas, G., Wang, Z., Yeager, M., Fraumeni, J.F., Jr., Skryabin, K.G., McKay, J.D., Rothman, N., Chanock, S.J., Lathrop, M. & Brennan, P. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet* **43**, 60-5 (2010).
22. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., Crenshaw, A., Cancel-Tassin, G., Staats, B.J., Wang, Z., Gonzalez-Bosquet, J., Fang, J., Deng, X., Berndt, S.I., Calle, E.E., Feigelson, H.S., Thun, M.J., Rodriguez, C., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F.R., Giovannucci, E., Willett, W.C., Cussenot, O., Valeri, A., Andriole, G.L., Crawford, E.D., Tucker, M., Gerhard, D.S., Fraumeni, J.F., Jr., Hoover, R., Hayes, R.B., Hunter, D.J. & Chanock, S.J. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* **40**, 310-5 (2008).
  23. Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C.S., Hughes, D., Warren-Perry, M., Tapper, W., Eccles, D., Evans, D.G., Hoening, M., Schutte, M., van den Ouweland, A., Houlston, R., Ross, G., Langford, C., Pharoah, P.D., Stratton, M.R., Dunning, A.M., Rahman, N. & Easton, D.F. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**, 504-7 (2010).
  24. Baird, D.M. Variation at the TERT locus and predisposition for cancer. *Expert Rev Mol Med* **12**, e16 (2011).
  25. Johnatty, S.E., Beesley, J., Chen, X., Macgregor, S., Duffy, D.L., Spurdle, A.B., deFazio, A., Gava, N., Webb, P.M., Rossing, M.A., Doherty, J.A., Goodman, M.T., Lurie, G., Thompson, P.J., Wilkens, L.R., Ness, R.B., Moysich, K.B., Chang-Claude, J., Wang-Gohrke, S., Cramer, D.W., Terry, K.L., Hankinson, S.E., Tworoger, S.S., Garcia-Closas, M., Yang, H., Lissowska, J., Chanock, S.J., Pharoah, P.D., Song, H., Whitmore, A.S., Pearce, C.L., Stram, D.O., Wu, A.H., Pike, M.C., Gayther, S.A., Ramus, S.J., Menon, U., Gentry-Maharaj, A., Anton-Culver, H., Ziogas, A., Hogdall, E., Kjaer, S.K., Hogdall, C., Berchuck, A., Schildkraut, J.M., Iversen, E.S., Moorman, P.G., Phelan, C.M., Sellers, T.A., Cunningham, J.M., Vierkant, R.A., Rider, D.N., Goode, E.L., Haviv, I. & Chenevix-Trench, G. Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and TERT, a cancer susceptibility "hot-spot". *PLoS Genet* **6**, e1001016 (2010).
  26. McKay, J.D., Hung, R.J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., McLaughlin, J., Shepherd, F., Montpetit, A., Narod, S., Krokan, H.E., Skorpén, F., Elvestad, M.B., Vatten, L., Njolstad, I., Axelsson, T., Chen, C., Goodman, G., Barnett, M., Loomis, M.M., Lubinski, J., Matyjasik, J., Lener, M., Oszutowska, D., Field, J., Liloglou, T., Xinarianos, G., Cassidy, A., Vineis, P., Clavel-Chapelon, F., Palli, D.,

- Tumino, R., Krogh, V., Panico, S., Gonzalez, C.A., Ramon Quiros, J., Martinez, C., Navarro, C., Ardanaz, E., Larranaga, N., Kham, K.T., Key, T., Bueno-de-Mesquita, H.B., Peeters, P.H., Trichopoulou, A., Linseisen, J., Boeing, H., Hallmans, G., Overvad, K., Tjonneland, A., Kumle, M., Riboli, E., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M. & Brennan, P. Lung cancer susceptibility locus at 5p15.33. *Nat Genet* **40**, 1404-6 (2008).
27. Shete, S., Hosking, F.J., Robertson, L.B., Dobbins, S.E., Sanson, M., Malmer, B., Simon, M., Marie, Y., Boisselier, B., Delattre, J.Y., Hoang-Xuan, K., El Hallani, S., Idbah, A., Zelenika, D., Andersson, U., Henriksson, R., Bergenheim, A.T., Feychting, M., Lonn, S., Ahlbom, A., Schramm, J., Linnebank, M., Hemminki, K., Kumar, R., Hepworth, S.J., Price, A., Armstrong, G., Liu, Y., Gu, X., Yu, R., Lau, C., Schoemaker, M., Muir, K., Swerdlow, A., Lathrop, M., Bondy, M. & Houlston, R.S. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* **41**, 899-904 (2009).

## **Appendices**

NA